Research Paper | Computer Science | E-ISSN No : 2454-9916 | Volume : 3 | Issue : 3 | Mar 2017

MEGA MART BI SOLUTION (RETAIL)

Shailesh P Bendale [1] | Harshvardhan Kadam [1] | Madhura Gajendragadkar [1] | Rohit Chougule [1] | Rohit Bapat [1]

[1] Department of Computer Engineering, NBN Sinhgad School of Engineering, Pune, India.

ABSTRACT

In recent times, technology applications in Business Intelligence (BI) have been developed rapidly. BI is considered to be one of the hottest emerging technologies. BI consists of a broad category of applications and technologies for gathering and storing data to analyze and help enterprises make smarter business decisions and strategies for designing and planning. For executive leaders, corporate managers, consultants, decision makers and business analysts, there is no longer a need to spend hours in designing and developing typical reports or charts, the entire solution can be implemented through using Business Intelligence "BI" software. The BI market is exciting, continually innovating and growing to meet the ever-expanding requirements of businesses of all magnitudes and sizes. This represents a vast competitive advantage. The current study uses the latest version of the main Open Source Business Intelligence suite: Jaspersoft for analysis and reporting both.

KEYWORDS: Business Intelligence, Jaspersoft, data integration, ETL, Reporting.

I. INTRODUCTION

Business intelligence (BI) is defined as the ability of an organization to take all its processes and capabilities and then convert these into knowledge, ultimately getting right information for the right people, at the right time, through the right channel. The aim of BI is to allow for easy analysis of huge dimensions of data and to categorize appropriate parts of this data which can be used for improving business strategies so as to generate more revenue. This can provide businesses with a competitive market gain and long-term solidity.

BI applications include the following:

- Data mining – for digging up useful and relevant data from a huge amount of data.
- Online analytical processing – for data analysis in warehouses.
- Analytics – for analyzing trends and patterns.
- Reporting – for representing data in a readable format.
- Business performance management – to increase future profits.
- Text mining- to search for specific keywords.

There are a huge number of tools that provide support to Business Intelligence solutions. These tools can be used for ETL, analysis or OLAP, Data mining, Database management and reporting purposes. Some of these tools are:

- ETL tools – IBM WebSphere DataStage, Teradata Parallel Transporter, SAS ETL Studio, Informatica PowerCenter etc.
- OLAP tools – Oracle Business Intelligence Discover 11g, SQL Server Analysis Services, DB2 OLAP Server, etc.
- Data mining tools – IBM Intelligent miner, Oracle Data miner, SAS Enterprise miner, Teradata warehouse miner, etc.
- Reporting tools – Jaspersoft Server, Cognos ReportNet, Crystal Reports Server, etc.
- Database management systems – Microsoft SQL Server, Oracle Database, DB2, Sybase IQ, etc.

The use of big data and business intelligence has already transformed business decision-making in global companies such as Amazon.com and Walmart.

Consider this example:

A store has multiple branches across many cities. It sells the same products everywhere for the same rate. During a monthly analysis, it was found that the sales of a particular product were higher in a certain branch as compared to all other branches. The store can then divert the stock of that product to this branch where more sales and therefore more profits, are being made. A store may also witness that the sales for a particular product are not as per the expectations, so this product can be put up on sale or some coupons can be made available to promote it. This is an application of business intelligence.

II. RELATED TOOLS:

A. Jaspersoft (Server and ETL):

Jaspersoft is known for providing a form of self-service tailored to the individual needs of companies. The Jaspersoft suite was created in 2006, after several years the company has created various tools individually. Jaspersoft provides the most flexible BI, which is economical and widely deployed in the world. The Jaspersoft website states that more than 14 million copies of open source software have been downloaded in the world, with 175,000 production deployments and over 14,000 customers in 100 countries. They also claim that the suite is updated frequently by a development community of more than 225,000 registered members. This community has the distribution of open source, free, and commercial distribution spread over three editions (Express, Professional and Enterprise). The distribution community is very limited as compared to commercial distributions and is distributed under a GNU GPL. It is composed of JasperReports Server modules, JasperReports Library, Jaspersoft ETL, Jaspersoft Studio and iReport Designer.

JasperReports Server is a standalone reporting server that can be embedded in any Java application. It provides reports and analysis that can be incorporated into Web applications or mobile applications. It also provides real-time or scheduled reports to the web, mobile, printer or e-mail in a variety of formats. It is optimized to share, protect and centrally manage the reports and analysis. Among the various features the ones that stand out are: formatting and interactively viewing reports, centralized and secure repository, generation, scheduling and distribution of reports and customizable interface. According to Jaspersoft, JasperReports Library is the most popular open source tool for creating reports. It is entirely programmed in Java and is able to use data from any source and produce documents which can be viewed, printed or exported in a variety of formats including HTML, PDF, Excel, OpenOffice and MSWord. Data integration (ETL - Extract, Transform, and Load) is supported by Jaspersoft ETL. This allows you to extract data from multiple sources, transform the data based on defined business rules and loads them into a data warehouse or data mart for analysis and reporting. Among the features include the graphical desktop environment, more than 500 connections to components and version control work.

The Jaspersoft Studio is a report design environment based on Eclipse for JasperReports and JasperReports Server. It lets you create reports from any data source, formatted for viewing on screen or print format, exporting to a wide range of formats. Among the features include the graphical desktop environment, the reporting models supported by themes, integration with JasperReports Server, sophisticated layouts with graphics, images, cross-sub-reports and tables, access data via JDBC, TableModels, JavaBeans , XML, Hibernate, CSV and custom backgrounds and publish the reports in PDF, RTF, XML, XLS, CSV, HTML, text files, DOCX or OpenOffice. Based on NetBeans, iReport Designer, Jaspersoft Studio is a tool to image, with essentially the same features. The latest version of Jaspersoft Studio is 5.0, the last update of December 2012.

B. HTML CSS and JavaScript:

HTML is a computer language specially used for creation of web pages or in general web sites. These websites are visible to each and every machine that is connected to the Internet and asks for the content of that particular website. It is comparatively very simple to learn as the basics are accessible to the majority people and quite influential in what it permits to create. With the growing requirements and demands of the Internet, HTML is constantly being refined and upgraded to meet the user requirements. The language is developed and maintained by W3C

and WHATWG, the organisations in charge of the maintenance and design of the language.

HyperText is used to move around on the web. It is done by clicking on special text called hyperlinks which links separate html documents or more precisely, web pages to each other. Hyper means that it is not linear, i.e. any number of web pages or documents can be viewed without and pre defined order by clicking on the hyperlinks provided on the web pages.

CSS stands for Cascading Style Sheets and is basically a style language and used for defining the layout of HTML documents. It basically deals with the fonts, margins, colors, background, lines, height, width, positions, and many more elements. CSS just provides a mechanism to format the structure content which is given by HTML. It provides several advanced and sophisticated techniques for enhanced and precise control over the layout from one single style sheet.

JavaScript is a scripting language. It is a lightweight programming language which is interpreted by the browser engine when it loads the webpage. It is basically used for various purposes such as putting text in an html page on-the-fly, making web pages responsive, detecting users' browsers, validating web form data, creating cookies, etc.

III. LITERATURE REVIEW

A significant amount of work has been carried out in this field and a lot of research has been dedicated towards making the Business Intelligence applications more user-friendly and interactive. A lot of emphasis is being given on making it suffice user needs.

Chung et al. [1] presented two detailed case studies on data integration and data mining which are essential steps for any Business Intelligence project. The first case study presented shows the conventional data analytics using relational database techniques like Oracle database and integrating and mining the data using Cognos BI tool. The second case study is intended towards multimedia data analytics using Mongo database and Pentaho BI tool for integration and mining of multimedia data obtainable on a company's website. Business Intelligence (BI) technologies offer historical, current and predictive analysis of business functions. BI technologies provide ad-hoc reporting, OLAP, business performance management, predictive analytics, competitive intelligence, decision making and benchmarking [2].

Business Intelligence permits any business or corporation's executives or decision makers to use the data on customer purchasing patterns, product profitability, demographics, and demand tendency to build successful strategic decisions. These decisions help the company plan its business and capitalize on profitability. Recommendation systems can be used for supporting better strategy and decision making for the executive officers. There are two kinds of applications of such systems: recommendation and prediction [3]. Collaborative filtering algorithm is used in these recommender systems. The similarity measure between users or items is used to recommend or predicate items that users have not noticed but might have interests in. [4]. This technology is ideally suitable for Internet-based e-commerce applications, as enterprise users can effortlessly obtain and analyze the preferences of users for products and recommend exact product to users with e-commerce. One application of BI system in microfinance institutions which is very similar to our proposed work can be seen in the work presented by Hajji et al. [5].

IV. METHODOLOGY

The methodology discussed in this section uses the concepts of Data integration and mining (Extract, Transform and Load), Online Analytic Processing (OLAP), Data warehousing, Reporting, Decision making, etc.

First of all, day-to-day transactions of each store need to be captured. These transactions are captured using OLTP systems such as MySQL/SQL. For the proposed work, MySQL database is considered as a suitable choice as it is open source. Every day transactions and sales are captured in the database which is defined using a fixed schema as devised by the team. Important tables in this schema being product table, store table, product category table, promotion table, point of sale table, coupon promotion table, etc. The sources of the data for implementation of data warehouse produce huge amount of data that needs to be reduced, transformed and finally loaded into the data warehouse.

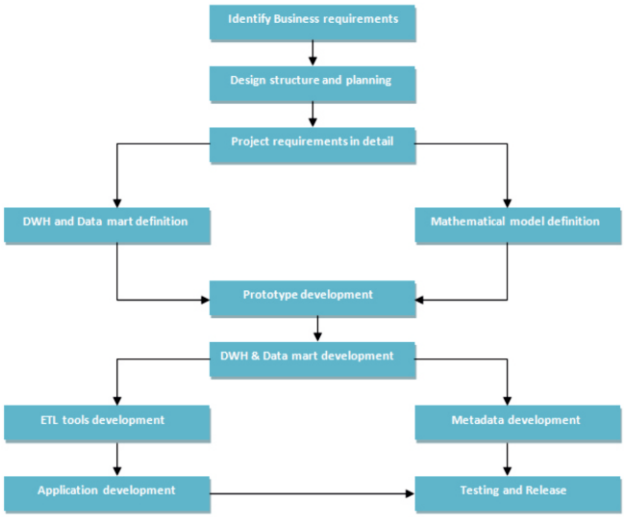

Fig 1: BI System Development Phases

Reducing the size of the data and finding only relevant data is an important task in the proposed project. Jaspersoft ETL tool is used for this purpose. It offers user friendly interface and also presents huge amounts of performance gains in terms of speed and accuracy. The extract phase of ETL fetches only relevant data from huge data sources with redundant data. This data may contain inconsistencies, missing values, naming inconsistencies, etc which might affect the performance of the BI system. Hence, these inconsistencies and erroneous values need to be corrected using various dedicated techniques. This comes under the transform phase. It also puts the corrected data in a correct format as decided by the team. The next and final step is to load this data into the warehouse on which analytical queries can be fired and useful information can be obtained. This information is then transformed into knowledge which is used to aid decision makers or executive officers to make better decisions based on the knowledge gained. For analysis of data in the data warehouse, Jaspersoft tool (plug-in in Netbeans) is used. It offers a number of features to the developers to create domains, create repositories, create ad-hoc views, manage users of the systems, generate reports, etc. Jaspersoft Community edition is freely available under the name JasperReports Server Community edition. The community edition of JasperReports Server can be used under the AGPL v3 license, which permits the use of this tool for free for our own use but does not allow it to be incorporated as a part of any commercial product. It simplifies the analysis task to a great extent.

The next and the most important part of this project is reporting. The end user of the proposed system can see the business growth or performance, or a product's sales, profitability, or even see how well the promotion on product is working, whether coupon promotion was successful or not, etc using simple graphical user interface. The UI of the proposed system is user friendly and easy to use. The user interface is interactive and consists of JSP and HTML pages with a customized layout designed using CSS. Each role is assigned a separate page with access permissions hard coded through the login database. The administrator has access to all pages and has special permissions to add and delete users. The generated reports can be viewed in the same page by using the 'iframe' concept. The reports are fetched from Jasper dynamically based on the option selected by the user. Filters are provided for the users to input their needs to the systems for better view of the area of interest. A simple dashboard is provided for this purpose and ad-hoc queries are fired at the back end to get the required data. Users can change the parameters as and when required and get the reports accordingly. The onclick event sends a predefined value to a JSP page which compares it to a hash map. The hash map consists of the report names associated with each value. The respective report is fetched dynamically and displayed as a PDF file which can be downloaded by the user. The interface also provides quick access to reports in the sidebar. The user has options like viewing all reports, viewing reports based on particular parameters, viewing graphical data generated from the report such as pie charts and bar graphs.

Reports can be fetched into a number of formats such as HTML or PDF. These reports are most important part of any BI project as they provide an insight into the business. The end user or executive officers can obtain useful trends and patterns with the help of these reports and act accordingly for betterment of the business.

V. FINDINGS

There are a number of observations and results that approve the above methodology. The suggested logic for implementation of the system can be realised by using MySQL, JasperReports Server, Jasper ETL, iReport designer for the database, ETL, OLAP and reporting respectively.

Alternative approach which could be used is to develop a web application using the methodology discussed in the previous section. Firstly, an official license of

Jaspersoft server will be needed to be purchased. Next, a Java Web Application can be developed that will run on the server machine of the domain and get input from the user and provide to the server. The server will then perform analysis based on the user inputs and provide reports as output.

VI. DISCUSSIONS AND LIMITATIONS

There are a variety of aspects of the software system which could be improved and enhanced. These aspects would help make a more reliable MegaMart BI system. Let us take a look at the possible limitations and challenges involved while realising this project.

A. Quality of data

The quality of the data available at hand from the various data sources cannot be used generously unless it is received from a trusted source.

B. Quantity of data

The amount of data available in the database determines the usability and correctness of the system. The size of this data is huge and needs to be reduced to extract only significant data.

C. Query optimization

The queries are complex and need to be optimized to enhance the system efficiency. Query plans can be considered to execute any given query efficiently.

D. Privacy and Security

The user credentials and privileges should be effectively maintained as private and kept secured. The secrecy of this data should not be hampered in any way.

Further, a few improvements can be done to the existing design. With the current methodology, reports are generated only when a user asks for it. In some cases, reports are required frequently and need to be scheduled without the user having to ask for it explicitly. Another improvement that can be done is to enhance the overall quality and efficiency of the web application and make it more interactive.

Another challenge would be to provide extension to the system for data based on different geographical regions and provide a system which would perform analysis on a vast geographical area.

VII. CONCLUSION

We have presented a possible solution to realise the vision of providing better understanding of the business to the MegaMart corporate. The methodology as discussed in the fourth section uses the concepts of Data integration and mining, data warehousing, online analytical processing and reporting to generate an efficient Business Intelligence solution. The technologies and tools put forward in section two can be used to develop this project. We have also discussed a few improvements and challenges in section five which will help in improving the scope of this project.

IV. REFERENCES

[1] P. Chung, S. Chung, “On Data Integration and Data Mining for developing Business Intelligence”, IEEE LISAT, 2013.

[2] P. Chung, S. Chung and C. Hui, “A Web Server Design using Search Engine Optimization Techniques for Web Intelligence for Small Organizations”, IEEE LISAT, 2012.

[3] Su, Xiaoyuan and Taghi M. Khoshgoftaar, “A Survey of collaborative filtering techniques” Advances in Artificial Intelligence, Vol. 2009, pp.4, 2009.

[4] Rajaraman, Anand and Jeffrey David Ullman, “Mining of massive datasets” Cambridge University Press, 2012.

[5] T. Hajji, S. Yasser el Jasouli, J. Mbarki, Jaara el Miloud, “Microfinance risk analysis using the Business Intelligence”, IEEE International Colloquium on Information Science and Technology (CIST), 2016.